2009

Situational Awareness Assessment in Flight Simulator Experiment

Henk van Dijk

Koen van de Merwe

Rolf Zon

SITUATIONAL AWARENESS ASSESSMENT IN FLIGHT SIMULATOR EXPERIMENT

Henk van Dijk, Koen van de Merwe and Rolf Zon
National Aerospace Laboratory NLR
Amsterdam, the Netherlands

Within the HILAS (Human Integration into the Lifecycle of Aviation Systems) project a flight simulator experiment was performed. The aim of the experiment was to study and select relevant Human Factors tools for pilot Situational Awareness assessment. One specific scenario was designed in which a malfunction of the aircraft was simulated: an Indicated Air Speed discrepancy. The malfunction was introduced during flight and slowly progressed over time while researchers monitored if and how pilots detected the discrepancy. Pilot behaviour was studied during the scenario; i.e. pilots' Situational Awareness was assessed via eye trackers and rating scales.

HILAS[1] stands for Human Integration into the Lifecycle of Aviation Systems. The objective of the Flight Deck Technologies (FDT) strand within the HILAS project is to create a Human Factors (HF) related set of tools that can be used to design and evaluate flight deck technologies. Such an HF set of tools can be applied as an instrument for HF certification. The importance of HF certification and its added value is explained by Jorna (2007) and McDonald (2007). The flight simulator experiment discussed in the current paper forms a part of the FDT strand.

Flight Simulator Experiment

The flight simulator experiment performed in the HILAS FDT strand comprises two phases. The results of the phase 1 experiment were previously discussed by Zon & Roerdink (2007). The current paper discusses part of the phase 2 experiment. The specific aim of the experiment discussed in the current paper was to study and select relevant HF tools for pilot Situational Awareness (SA) assessment.

Participating pilots

Six crews of each two airline pilots (i.e. a captain and a first officer) participated in the experiment. There are two pilot tasks in the simulation: pilot flying (PF) and pilot not flying (PNF). The tasks to be carried out by PF and PNF match normal operations and are varied between captain and first officer.

Flight simulator

GRACE (Generic Research Aircraft Cockpit Environment) is a generic flight simulator, representing a modern large two-engine fly-by-wire airliner (see GRACE cockpit in Figure 1). GRACE has a number of standard configurations. For the current experiment the Airbus A320 configuration was selected. A high fidelity simulator such as GRACE allows researchers to perform realistic experiments in a fully controlled environment.

---

[1] The Human Integration into the Lifecycle of Aviation Systems (HILAS) project is part of the 6th framework programme for aeronautics and space research, sponsored by the European Commission. The overall aim of the HILAS project is to develop a model of good practice for the integration of human factors across the life-cycle of aviation systems. The project contains four strands of work (see for further information http://www.hilas.info/mambo/).

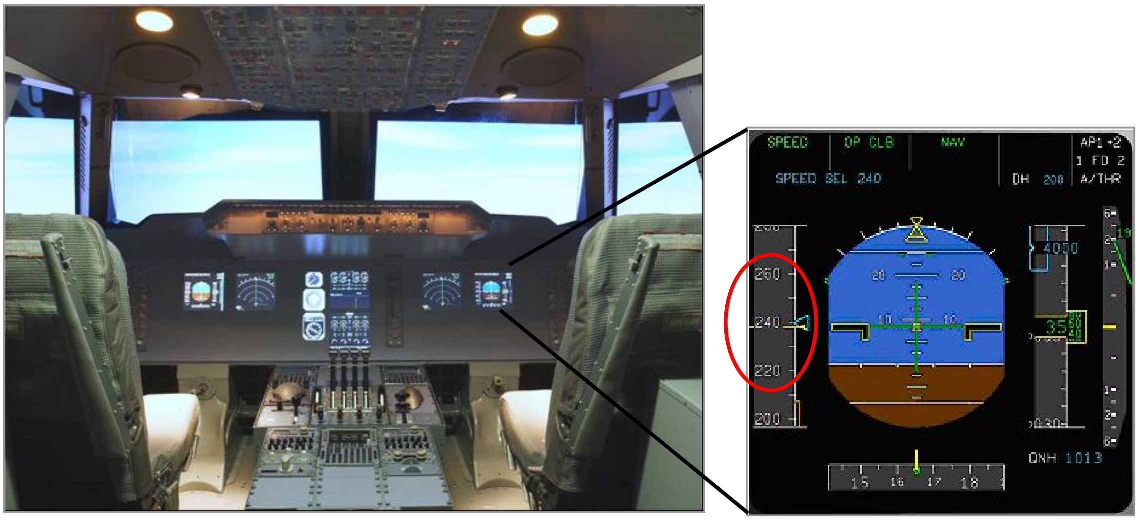

Figure 1. GRACE cockpit on the left side and PFD on the right side. The speed tape on the PFD is circled in red.

Experimental scenarios

The flight consisted of a trip from London Heathrow to Amsterdam Schiphol (starting in cruise, ending in landing). A specific scenario was designed in which SA was hampered. This was done by simulating a malfunction of the aircraft during the flight: an Indicated Air Speed (IAS) discrepancy was introduced. The discrepancy was indicated by the two available Primary Flight Displays (PFDs); i.e. one display showed the correct air speed while the other showed a lower, false air speed (see Figure 1 for a PFD with the speed tape circled in red). Once the discrepancy was initiated by the simulator after about 10 minutes in flight, it slowly progressed over time while researchers monitored if and how the pilots detected the discrepancy, and if and how the pilots figured out what the correct air speed was. As far as the crew concerns, they were flying a normal flight until the malfunction was detected. The flight duration of this scenario was 25 minutes.

It was expected –as the discrepancy progressed over time– that most of the crews detected the malfunction earlier than after the 2 minutes on which an engine display warming was given. As it turned out, none of the six crews discovered the specific discrepancy on both PFDs before the engine display indicated the air speed discrepancy. The crews were informed of the malfunction with an auditory warning after 2 minutes. Consequently, the analysis focussed only on the time that the crew needed to figure out the correct air speed (i.e. after the engine display warning), and not on the time they needed to detect the malfunction. The period after the warning until the moment of detection of the malfunction is referred to as post-period. In the analysis, this post-period is compared to a reference period that has the same length as the post-period and takes place immediately before the onset of the IAS discrepancy (see Figure 2 for an illustration of this time-line).

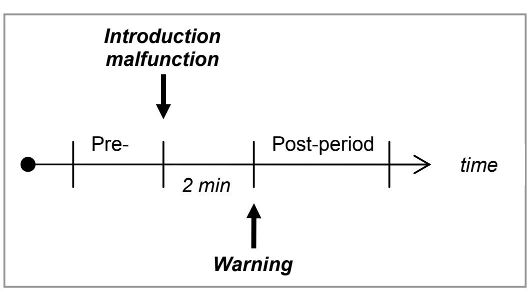

Figure 2. Illustration of the time-line for the analysis.

HF tools

During this SA scenario, pilot behaviour was studied using the following HF tools:

Eye tracking

The eye tracker that was used in the current experiment is the Applied Science Laboratories (ASL[2]) 6000 with Ascension Technologies optical head tracker. Two trackers were placed in the cockpit (one for each pilot). The pilots wore a headband on which the optronics were mounted.

Dwell time on the PFDs (crewmember's own PFD and cross check) and entropy were used as eye tracking measures for the PF and PNF. Dwell time provides information regarding the amount of time spent viewing the PFD and can be interpreted as a measure of attention. Entropy provides information regarding the search strategies used by the crewmembers and can be interpreted as a measure of randomness of the viewing pattern.

ISA

The simple rating technique called Instantaneous Self Assessment (ISA) was used in the current scenario to measure the pilots' overview of the situation on that particular moment in the scenario; i.e. the ratings were self assessed during the flight. The pilot was asked (every other 2 minutes) to respond to the rating scale presented on the touch screen display (an electronic flight bag) placed in front of him/her by assessing his/her current situation overview (5 being very high and 1 being very low; see ISA rating scale in Figure 3).

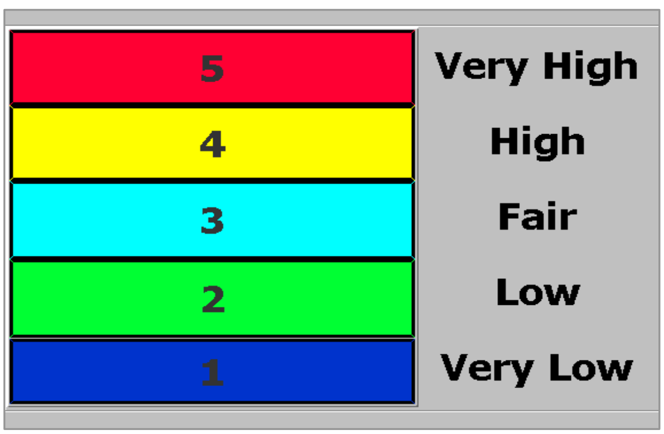

Figure 3. ISA rating scale displayed on an electronic flight bag.

Results

An α of 5% was used for significance testing and Cohen's *d* was used as a measure of effect size.

Eye tracking

The difference in viewing behaviour between the period in which the crew was looking for the error (post-period) and the reference period before the error was introduced (pre-period) was analysed. A paired-samples t-test showed significantly longer time spent dwelling on the crewmember's own PFD in the post-period compared to the pre-period ($t(9) = -2.326$, $p < .05$, $d = -.74$). Similar results were found for cross check behaviour ($t(9) = -4.005$, $p < .01$, $d = -1.27$) and all PFDs ($t(9) = -2.916$, $p < .05$, $d = -.92$).

[2] See for further information http://www.a-s-l.com/.

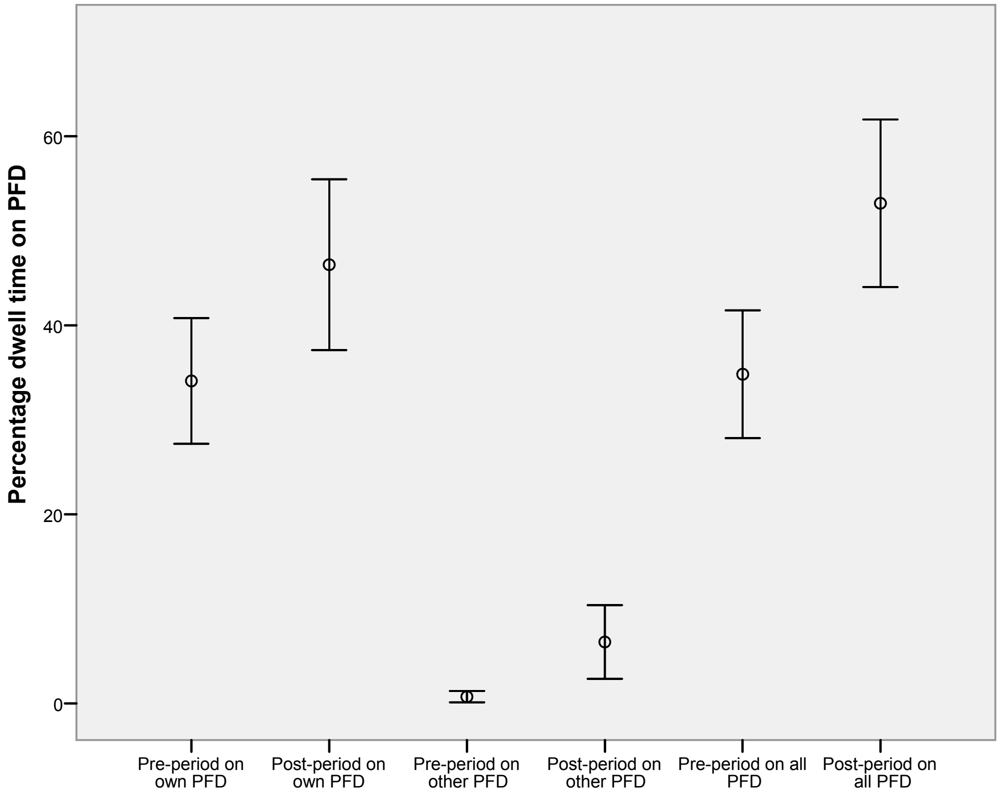

Figure 4. Percentage dwell time on the crewmember's own, other and all PFDs in the pre-period vs. the post-period. All differences between the pre- and the post-period were statistically significant.

The relationship between the time it took for the crew to discover the IAS discrepancy on the PFDs and the amount of time spent looking at the different PFDs was also analysed. It was assumed that the malfunction could only be discovered by cross checking both PFDs and comparing the information presented on them. A significant negative correlation was found between the discovery period and the amount of time spent cross checking the other PFD ($r(12) = -.613, p < .05$). This means that the more time the crewmember spent on cross checking the other PFD, the less time it took to discover the IAS discrepancy. There was no significant correlation found for the time spent on one's own PFD.

The scanning patterns of the crews were investigated using a t-test. The results showed a significant increase in entropy during the post-period ($t(10) = -2.347$, $p < .05$, $d = -.71$). This means that the patterns followed by the eye movements during the post-period were less systematic than during the pre-period indicated searching behaviour.

ISA

The average of multiple ISA answers per crew were investigated using a t-test ($n = 28$). ISA ratings were significantly lower in the post-period than in the pre-period ($t(27) = 2.780$, $p < .05$. $d = .52$). This means that the crew reported that they experienced a reduced SA in the post-period.

Key Findings

The current paper discussed the phase 2 high fidelity simulator experiment within the HILAS FDT strand. The experiment studied and selected relevant HF tools for pilot SA assessment. A specific SA scenario was designed in which a simulated malfunction was introduced during the flight; i.e. after a certain amount of time in flight, an IAS discrepancy was set on and slowly progressed over time.

Eye measures comprise a number of variables that can be measured and recorded with an eye tracker. The measures can be used as indicators of fatigue, mental or visual workload and also as indicators of attention. The eye tracking measures in the current experiment revealed the eye scanning patterns; i.e. where did pilots focus on during the period in which they tried to discover the correct air speed and how quickly did they make the discovery. As such, and supplemented by the ISA ratings, this informed the researchers about the perceived SA of the pilots during the course of the flight.

Eye tracking

Eye tracking can be considered as an indirect measurement of pilot attention and focus on tasks. The basic assumption is that looking at a certain location means that attention is focussed on this particular location. The analyses of the eye tracker data primarily focussed on the two PFD locations since these locations were the designated displays on which the IAS discrepancy was revealed. Of course, this is a simplification of the reality as there are other methods in finding out the correct air speed (e.g. checking with air traffic controller, looking at ground speed on navigation display).

The dwell time results indicate an increase of dwell time after the engine display warning on the pilot's own PFD and the other's PFD (cross check). This was as expected since the warning of an air speed discrepancy on the engine display encourages in looking at the different PFDs (as air speed is indicated here). Interestingly, the dwell time on the other's PFD correlates negatively with the time it takes to discover the correct air speed (the duration of the post-period decreases when the time used for the cross check increases). This result corresponds with the notion that the cross check is evident in discovering the correct air speed.

The entropy results reveal a higher entropy value in the post-period compared to the pre-period. This was as hypothesized. Random screening after the engine display warning increases because the pilots are searching for the solution.

ISA

Self rating techniques such as ISA are frequently used to elicit subjective estimates of SA from participants. These rating scales are typically administered post-flight. The primary advantage of such techniques is their low cost, ease of implementation and non-intrusive nature. However, self rating techniques administered post-flight suffer from a number of disadvantages that are associated with reporting SA data "after the fact". These include that participants are prone to "forgetting" periods of the flight when they possessed a poor level of SA, and more readily remember the periods when they possessed a superior level of SA (Endsley, 1995). Therefore, in the HILAS phase 2 experiment, the ISA technique was implemented in such a manner that ratings could be assessed during the course of the flight; i.e. pilots could rate their SA on a particular given moment in the scenario. This helps to pinpoint the ups and downs in the SA of the specific pilot. As it turns out, assessing ISA in flight using an electronic flight bag is an easy-to-implement and easy-to-use technique in an experiment. The crew did not report any problems (e.g. distraction of the flying task or intrusiveness of the tool) and rated their SA immediately after the scale popped up on the electronic flight bag.

The ISA results (for PF and PNF) show a decrease in SA after the engine display warning. This was as expected since this particular SA scenario is set up to hamper SA.

Summarizing, crew behaviour –specifically the eye scanning patterns– has been proven to change due to the compromised SA. This is clearly indicated by the results of the eye tracking measures dwell time and entropy. The ISA results add to this the perceived SA of the crew during the course of the flight. Together, eye tracking and ISA provide a more complete picture than both measures independently.

Acknowledgements

The authors would like to thank the HILAS FDT strand members for their contribution in the experiment. Further, we would like to thank the European Commission for sponsoring this research.

References

Jorna, P. G. A. M. (2007). Human performance enhancements: from certification to HCI innovation. In: D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, HCII 2007, LNAI 4562* (pp. 697-704). Berlin Heidelberg: Springer-Verlag.

McDonald, N. (2007). Human integration in the lifecycle of aviation systems. In: D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, HCII 2007, LNAI 4562* (pp. 760-769). Berlin Heidelberg: Springer-Verlag.

Zon, G. D. R., & Roerdink, M. I. (2007). HCI testing in flight simulator: set up and crew briefing procedures. Design and test cycles for the future. In: D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, HCII 2007, LNAI 4562* (pp. 867-876). Berlin Heidelberg: Springer-Verlag.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37,* 65-84.